

OIPE
~~Deny~~

RAW SEQUENCE LISTING
PATENT APPLICATION: US/09/785,269

DATE: 04/26/2001
TIME: 10:29:41

Input Set : N:\CrF3\RULE60\09785269.txt
Output Set: N:\CRF3\04262001\I785269.raw

ENTERED

SEQUENCE LISTING

4 (1) GENERAL INFORMATION:

6 (i) APPLICANT: IMAI, Kensaku
7 KITAJIMA, Masato

9 (ii) TITLE OF INVENTION: METHOD AND APPARATUS FOR AUTOMATICALLY
10 REMOVING VECTOR UNIT IN DNA BASE SEQUENCE

12 (iii) NUMBER OF SEQUENCES: 19

14 (iv) CORRESPONDENCE ADDRESS:

15 (A) ADDRESSEE: Staas & Halsey
16 (B) STREET: 700 Eleventh Street, N.W., Suite 500
17 (C) CITY: Washington
18 (D) STATE: DC
19 (E) COUNTRY: US
20 (F) ZIP: 20001

22 (v) COMPUTER READABLE FORM:

23 (A) MEDIUM TYPE: Floppy disk
24 (B) COMPUTER: IBM PC compatible
25 (C) OPERATING SYSTEM: PC-DOS/MS-DOS
26 (D) SOFTWARE: Patentin Release #1.0, Version #1.30

28 (vi) CURRENT APPLICATION DATA:

29 (A) APPLICATION NUMBER: US/09/785,269
30 (B) FILING DATE: 20-Feb-2001
31 (C) CLASSIFICATION:

34 (vii) PRIOR APPLICATION DATA:

35 (A) APPLICATION NUMBER: US 08/684,674
36 (B) FILING DATE: 22-JUL-1996

39 (viii) ATTORNEY/AGENT INFORMATION:

40 (A) NAME: Herbert, William F.
41 (B) REGISTRATION NUMBER: 31,024
42 (C) REFERENCE/DOCKET NUMBER: 862.1335/WFH

44 (ix) TELECOMMUNICATION INFORMATION:

45 (A) TELEPHONE: 2024341500
46 (B) TELEFAX: 2024341501

49 (2) INFORMATION FOR SEQ ID NO: 1:

51 (i) SEQUENCE CHARACTERISTICS:

52 (A) LENGTH: 57
53 (B) TYPE: nucleic acid
54 (C) STRANDEDNESS: double
55 (D) TOPOLOGY: linear

57 (ii) MOLECULE TYPE: DNA (genomic)

62 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

64 AAGCTTGCAT GCCTGCAGGT CGACTCTAGA GGATCCCCGG GTACCGAGCT CGAATTC

57

66 (2) INFORMATION FOR SEQ ID NO: 2:

68 (i) SEQUENCE CHARACTERISTICS:

69 (A) LENGTH: 18
70 (B) TYPE: nucleic acid
71 (C) STRANDEDNESS: double

RAW SEQUENCE LISTING
PATENT APPLICATION: US/09/785,269

DATE: 04/26/2001
TIME: 10:29:41

Input Set : N:\CrF3\RULE60\09785269.txt
Output Set: N:\CRF3\04262001\I785269.raw

72 (D) TOPOLOGY: linear
74 (ii) MOLECULE TYPE: DNA (genomic)
79 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:
81 TGCACTTGAA CGCATGCT 18
83 (2) INFORMATION FOR SEQ ID NO: 3:
85 (i) SEQUENCE CHARACTERISTICS:
86 (A) LENGTH: 17
87 (B) TYPE: nucleic acid
88 (C) STRANDEDNESS: double
89 (D) TOPOLOGY: linear
91 (ii) MOLECULE TYPE: DNA (genomic)
96 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:
98 TGCACTTGAA CGCTGCT 17
100 (2) INFORMATION FOR SEQ ID NO: 4:
102 (i) SEQUENCE CHARACTERISTICS:
103 (A) LENGTH: 17
104 (B) TYPE: nucleic acid
105 (C) STRANDEDNESS: double
106 (D) TOPOLOGY: linear
108 (ii) MOLECULE TYPE: DNA (genomic)
113 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4: 17
115 TGCACTTGAC GCATGCT
117 (2) INFORMATION FOR SEQ ID NO: 5:
119 (i) SEQUENCE CHARACTERISTICS:
120 (A) LENGTH: 17
121 (B) TYPE: nucleic acid
122 (C) STRANDEDNESS: double
123 (D) TOPOLOGY: linear
125 (ii) MOLECULE TYPE: DNA (genomic)
130 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:
132 TGCACTTGAC GCAAGCT 17
134 (2) INFORMATION FOR SEQ ID NO: 6:
136 (i) SEQUENCE CHARACTERISTICS:
137 (A) LENGTH: 17
138 (B) TYPE: nucleic acid
139 (C) STRANDEDNESS: double
140 (D) TOPOLOGY: linear
142 (ii) MOLECULE TYPE: DNA (genomic)
147 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:
149 TGCCTTGAAC GCATGCT 17
151 (2) INFORMATION FOR SEQ ID NO: 7:
153 (i) SEQUENCE CHARACTERISTICS:
154 (A) LENGTH: 2686
155 (B) TYPE: nucleic acid
156 (C) STRANDEDNESS: double
157 (D) TOPOLOGY: linear
159 (ii) MOLECULE TYPE: DNA (genomic)
164 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:
166 TCGCGCGTTT CGGTGATGAC GGTGAAAACC TCTGACACAT GCAGCTCCCG GAGACGGTCA 60

RAW SEQUENCE LISTING
PATENT APPLICATION: US/09/785,269

DATE: 04/26/2001
TIME: 10:29:41

Input Set : N:\CrF3\RULE60\09785269.txt
Output Set: N:\CRF3\04262001\I1785269.raw

168	CAGCTTGCTC	GTAAGCGGAT	GCCGGGAGCA	GACAAAGCCC	TCAAGGGCCG	TCAGGGGGTG	120
170	TTGGGGGGTC	TCGGGGTGG	CTTAACTATG	CGGCATCAGA	CGAGATTGTA	CTGAGACTGC	180
172	ACCATATGCG	GTGTTGAAATA	CGGCCACAGAT	GGCTAAGGAG	AAAATACCGC	ATCAGGGGCC	240
174	ATTCGCCATT	CAGGCTCGGC	AACCTGTTGGG	AAAGGGGATC	GTTGCGGGCC	TCTTCGCTAT	300
176	TACGCCAGT	GGCAGAACGG	GGATGTGCTG	CAAGGCATT	AAGTTGGTA	ACGGCAGGGT	360
178	TTTCCAGCT	ACGACGTTGC	AAAACGACGG	CGCATGCTTAC	GCTTCGATGC	CTGCAGGTGCG	420
180	ACTCTAGAGG	ATCCCCGGGT	ACCGAGCTGC	AATTCTTAT	CTATGGTCTA	GCTGTTCT	480
182	GTGTTGAATT	GTATCCCGT	CAACATTCGA	CACAAACATAC	GAGCCGAAG	CATAAAAGTGT	540
184	AAAGCCTGGG	GTGCTTAATG	AGTGAGCTAA	CTACATTAA	TTGGCTTGC	CTCACGCC	600
186	GCTTCCAGT	CGGGAAACCT	GTCTGCCAG	CTGCATTAT	GAATCGGCCA	ACGCGGGGG	660
188	AGAGCGCGTT	TGCGTATTGG	GGCTCTTCC	GCTCTTCG	CTACTGACTC	GCTGCGCTCG	720
190	GTCTGTCGGG	TGCGCGACG	GGTATCAGCT	CACTTAAAGG	CGGTAAATCC	GTATCCACA	780
192	GAATCAGGGG	ATAACGCAAGG	TAACAGCATG	TGACCAAAGG	GCCACAAAAA	GCCAGGAAC	840
194	CGTAAAAAGG	CGCGGTGTTG	GGCGTTTTTC	CATAGGCTC	CCCCCTTGA	CGAGCATCAC	900
196	AAAAATCGAC	GCTCAAGTC	GAGGTGGCGA	ACCCGACAG	GACTATAAG	ATACCAAGGC	960
198	TTGTTGAAGG	GAAGCTCCCT	CTGTGCGCT	CTCTGTTCC	CCCTGCGCT	TACCGGATAC	1020
200	CTGTCGGCT	TTCTCCCTTC	GGGAAAGCTG	GGCGCTTCTC	AAAGCTTACCO	CTGAGGTAT	1080
202	CTCAGTCGG	TGAGGTGTTG	TCGCTCCAG	CTGGCTGTC	TGCAAGGAAC	CCCCGCTCAG	1140
204	CCCGACCGCT	GGGCTTATTC	CGGTAACTAC	CGCTTGAAGG	CCAAACCGGT	AAAGACAGAC	1200
206	TTATGCCAAC	TGCGCAGGAC	CACTGTTAAC	AGATTAGTCA	GAGCGAGGTA	TGAGGGCGT	1260
208	GCTACAGAGT	CTTGTGAAGT	GTGCGCTAAC	TCAGGCTTACA	CTAGAGAAC	ACTATTGCT	1320
210	ATCIGCCTC	TGCTGAGGCC	AGTTACCTTC	GGAAAAGAG	TTGGTAGCTC	TTGATCCGC	1380
212	AAACAAACCA	CCGGCTTCTG	CGGTGTTTTT	TTTGTGTC	AGCACGAGAT	TACGCCGAGA	1440
214	AAAAAAGAT	CTCAAGAAGA	TCTCTTGTAT	TTTCTCTACCG	GCTCTGACGC	CTAGTGGAAAC	1500
216	GAAGAACTAC	GTTAAGGGAT	TTTGTGTCAG	AGATTATCA	AAAGGATCTT	CACCTAGATC	1560
218	CTTTAAATT	AAAAATGAA	TTTAAATAC	ATCTAAAGTA	TATATGAGTA	AACTTGTCT	1620
220	GACAGTTCAC	ATGCTTTAAT	CAGTGGGCA	CCATTCTTCG	CGATGTTGCT	ATTTCGTTCA	1680
222	TCATCATGTT	CTCTGAGTC	CTCTGTTGAT	AACTACTAGC	TACGGGAGGC	CTTACCATCT	1740
224	GGCCCGCATG	TCGAATGAT	ACCCGGAGAC	CCACCTCTAC	CGGCTCCAGA	TTTATCAGCA	1800
226	ATAAACACG	CAGCCGGGAC	GGCCGGAGGC	AGAAAGTGGTC	CTGCAACTTT	ATCCGCTCCC	1860
228	ATCCAGCTCA	TTAATTGTTG	CCGGGAGGT	AGAGTAAGTA	GTTCGCGCT	TAATAGTTG	1920
230	CCGCAACCTTG	TTGGCATTCG	TAACGGCATC	GTGTTGTCAG	GCTCTCGTT	TGTTATGGCT	1980
232	TCATTCAGCT	CGGGTCTCCA	ACAGCAAGG	CGAGTTACAT	GATCCCCCAT	GTGTTGCAA	2040
234	AAAGCGGTTA	GCTCTTCGG	TCCTCCGATC	GTTGTCAGAA	GTAAAGTTGC	CCAGGTGTTA	2100
236	TCACCTATGG	TTATGGCAGC	ACTGCATATA	TCTCTTACTC	TCATGCCATC	CTAAAGATGC	2160
238	TTTTCGTTGA	CTGGTAGTGA	CTCAACCAGC	TCATCTTGAC	AAATAGTGTAT	CGCCGCAGCC	2220
240	AGTTGCTCTT	CGCCGGCTTC	AATAACGGAT	AATAACGGGC	CACATAGCGA	AACTTTAAAAA	2280
242	GGCTCATCATCA	TTGGAAAAGC	TTCTTGGGG	CGAAAACCTG	CGAGGATCTT	ACCCGCTGTT	2340
244	AGATCCAGT	CGATGTAACC	CACTCGTGC	CCCAACTGAT	CTTCAGCATC	TTTACTTT	2400
246	ACCAAGCTTT	CTGGGTGAGC	AAAACAGGA	AGGCAAAATG	CGGCAAAAAA	GGGAATAAGG	2460
248	GGCACACGGA	AAATTGGAAT	ACTCTACTAC	TTCTTTTTC	AAATATTATG	AAGCATTT	2520
250	CAGGGTTATT	GTCTCATGAG	CGGATACATA	TTTGAATGTA	TTTAAAGAAA	TTAAACAAATA	2580
252	GGGGTCCCG	GCACATTTCC	CCGAAAGATG	CCACCTGAGC	TCTAAAGAAC	CATTATTATC	2640
254	ATGACATTAAC	CCTATAAAAA	TAGGGCTATAC	ACCGAGGCC	TTTCGTC		2686

256 (2) INFORMATION FOR SEQ ID NO: 8:

- 258 (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 66
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double

RAW SEQUENCE LISTING DATE: 04/26/2001
PATENT APPLICATION: US/09/785,269 TIME: 10:29:41

Input Set : N:\Crfr3\RULE60\09785269.txt
Output Set: N:\CRFR3\04262001\I785269.raw

262 (D) TOPOLOGY: linear
264 (ii) MOLECULE TYPE: DNA (genomic)
269 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:
271 GTGCCAACTG TGCATGCTG CAGGTGACT CTAGAGGATC CCCGGTACCG AGCTCGAATT
273 CCTATAA
275 (2) INFORMATION FOR SEQ ID NO: 9:
277 (i) SEQUENCE CHARACTERISTICS:
278 (A) LENGTH: 6
279 (B) TYPE: nucleic acid
280 (C) STRANDEDNESS: double
281 (D) TOPOLOGY: linear
283 (ii) MOLECULE TYPE: DNA (genomic)
288 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:
290 AAGCTT
292 (2) INFORMATION FOR SEQ ID NO: 10:
294 (i) SEQUENCE CHARACTERISTICS:
295 (A) LENGTH: 6
296 (B) TYPE: nucleic acid
297 (C) STRANDEDNESS: double
298 (D) TOPOLOGY: linear
300 (ii) MOLECULE TYPE: DNA (genomic)
305 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:
307 GCATGC
309 (2) INFORMATION FOR SEQ ID NO: 11:
311 (i) SEQUENCE CHARACTERISTICS:
312 (A) LENGTH: 6
313 (B) TYPE: nucleic acid
314 (C) STRANDEDNESS: double
315 (D) TOPOLOGY: linear
317 (ii) MOLECULE TYPE: DNA (genomic)
322 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:
324 CTGCAG
326 (2) INFORMATION FOR SEQ ID NO: 12:
328 (i) SEQUENCE CHARACTERISTICS:
329 (A) LENGTH: 6
330 (B) TYPE: nucleic acid
331 (C) STRANDEDNESS: double
332 (D) TOPOLOGY: linear
334 (ii) MOLECULE TYPE: DNA (genomic)
339 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:
341 GGTCAC
343 (2) INFORMATION FOR SEQ ID NO: 13:
345 (i) SEQUENCE CHARACTERISTICS:
346 (A) LENGTH: 6
347 (B) TYPE: nucleic acid
348 (C) STRANDEDNESS: double
349 (D) TOPOLOGY: linear
351 (ii) MOLECULE TYPE: DNA (genomic)
356 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

RAW SEQUENCE LISTING
PATENT APPLICATION: US/09/785,269

DATE: 04/26/2001
TIME: 10:29:41

Input Set : N:\Crf3\RULE60\09785269.txt
Output Set: N:\CRF3\04262001\I785269.raw

358 TCTAGA
360 (2) INFORMATION FOR SEQ ID NO: 14:
362 (i) SEQUENCE CHARACTERISTICS:
363 (A) LENGTH: 6
364 (B) TYPE: nucleic acid
365 (C) STRANDEDNESS: double
366 (D) TOPOLOGY: linear
368 (ii) MOLECULE TYPE: DNA (genomic)
373 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:
375 GTCGAC 6
377 (2) INFORMATION FOR SEQ ID NO: 15:
379 (i) SEQUENCE CHARACTERISTICS:
380 (A) LENGTH: 6
381 (B) TYPE: nucleic acid
382 (C) STRANDEDNESS: double
383 (D) TOPOLOGY: linear
385 (ii) MOLECULE TYPE: DNA (genomic)
390 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:
392 GTCCGAC 6
394 (2) INFORMATION FOR SEQ ID NO: 16:
396 (i) SEQUENCE CHARACTERISTICS:
397 (A) LENGTH: 6
398 (B) TYPE: nucleic acid
399 (C) STRANDEDNESS: double
400 (D) TOPOLOGY: linear
402 (ii) MOLECULE TYPE: DNA (genomic)
407 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:
409 CCCGGG 6
411 (2) INFORMATION FOR SEQ ID NO: 17:
413 (i) SEQUENCE CHARACTERISTICS:
414 (A) LENGTH: 6
415 (B) TYPE: nucleic acid
416 (C) STRANDEDNESS: double
417 (D) TOPOLOGY: linear
419 (ii) MOLECULE TYPE: DNA (genomic)
424 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:
426 GAATTG 6
428 (2) INFORMATION FOR SEQ ID NO: 18:
430 (i) SEQUENCE CHARACTERISTICS:
431 (A) LENGTH: 6
432 (B) TYPE: nucleic acid
433 (C) STRANDEDNESS: double
434 (D) TOPOLOGY: linear
436 (ii) MOLECULE TYPE: DNA (genomic)
441 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:
443 CCCGGG 6
445 (2) INFORMATION FOR SEQ ID NO: 19:
447 (i) SEQUENCE CHARACTERISTICS:
448 (A) LENGTH: 6

VERIFICATION SUMMARY

PATENT APPLICATION: US/09/785,269

DATE: 04/26/2001

TIME: 10:29:42

Input Set : N:\Crf3\RULE60\09785269.txt

Output Set: N:\CRF3\04262001\I785269.raw

L:29 M:220 C: Keyword misspelled or invalid format, [(A) APPLICATION NUMBER:]

L:30 M:220 C: Keyword misspelled or invalid format, [(B) FILING DATE:]